PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : G01N 33/53 — A1

(11) International Publication Number: **WO 96/24847**

(43) International Publication Date: 15 August 1996 (15.08.96)

(21) International Application Number: PCT/US96/02490

(22) International Filing Date: 12 February 1996 (12.02.96)

(30) Priority Data:

08/386,381	10 February 1995 (10.02.95)	US
08/400,220	7 March 1995 (07.03.95)	US
08/497,357	30 June 1995 (30.06.95)	US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/497,357 (CIP)
Filed on 30 June 1995 (30.06.95)

(71) Applicant *(for all designated States except US)*: SMITHKLINE BEECHAM CORPORATION [US/US]; Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

(72) Inventor; and
(75) Inventor/Applicant *(for US only)*: DUNNINGTON, Damien, John [GB/US]; 23 Forsythia Court, Lafayette Hill, PA 19444 (US).

(74) Agents: DUSTMAN, Wayne, J. et al.; SmithKline Beecham Corporation, Corporate Intellectual Property, UW2220, 709 Swedeland Road, P.O. Box 1539, King of Prussia, PA 19406-0939 (US).

(81) Designated States: JP, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: A PROCESS FOR IDENTIFIYING PHARMACEUTICALLY ACTIVE AGENTS USING AN EPITOPE-TAGGED LIBRARY

(57) Abstract

A process for indentifying pharmaceutically active agents is disclosed, which comprises simultaneously expressing a plurality of uniquely tagged target agents to form a target library, preparing an agent candidate pool and testing the target library and the agent candidate pool in an assay which identifies agents having desired characteristics. Also disclosed is a process for simultaneously expressing a plurality of uniquely tagged genes/gene products or membrane receptors to form a target library. In addition, a target library comprising a plurality of expressed uniquely tagged genes/gene products or membrane-bound receptors is disclosed, as are spatially encoded target libraries and methods of screening such libraries for pharmaceutical activity.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	**GB**	United Kingdom	**MW**	Malawi
AT	Austria	**GE**	Georgia	**MX**	Mexico
AU	Australia	**GN**	Guinea	**NE**	Niger
BB	Barbados	**GR**	Greece	**NL**	Netherlands
BE	Belgium	**HU**	Hungary	**NO**	Norway
BF	Burkina Faso	**IE**	Ireland	**NZ**	New Zealand
BG	Bulgaria	**IT**	Italy	**PL**	Poland
BJ	Benin	**JP**	Japan	**PT**	Portugal
BR	Brazil	**KE**	Kenya	**RO**	Romania
BY	Belarus	**KG**	Kyrgystan	**RU**	Russian Federation
CA	Canada	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CF	Central African Republic			**SE**	Sweden
CG	Congo	**KR**	Republic of Korea	**SG**	Singapore
CH	Switzerland	**KZ**	Kazakhstan	**SI**	Slovenia
CI	Côte d'Ivoire	**LI**	Liechtenstein	**SK**	Slovakia
CM	Cameroon	**LK**	Sri Lanka	**SN**	Senegal
CN	China	**LR**	Liberia	**SZ**	Swaziland
CS	Czechoslovakia	**LT**	Lithuania	**TD**	Chad
CZ	Czech Republic	**LU**	Luxembourg	**TG**	Togo
DE	Germany	**LV**	Latvia	**TJ**	Tajikistan
DK	Denmark	**MC**	Monaco	**TT**	Trinidad and Tobago
EE	Estonia	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**UG**	Uganda
FI	Finland	**ML**	Mali	**US**	United States of America
FR	France	**MN**	Mongolia	**UZ**	Uzbekistan
GA	Gabon	**MR**	Mauritania	**VN**	Viet Nam

A PROCESS FOR IDENTIFYING PHARMACEUTICALLY ACTIVE AGENTS USING AN EPITOPE-TAGGED LIBRARY

FIELD OF THE INVENTION

The present invention relates to an improved process for identifying pharmaceutically active agents.

BACKGROUND OF THE INVENTION

Over the past ten years, there has been a growing demand for the production and identification of agents that have pharmacological activity as, for example, agonists or antagonists of various cellular acceptor molecules, such as cell-surface receptors, enzymes or antibodies. In the continuing search for new chemical moieties that can effectively modulate a variety of biological processes, the standard method for conducting a search is to screen a variety of pre-existing chemical moieties, for example, naturally occurring compounds or compounds which exist in synthetic libraries or databanks. The biological activity of the pre-existing chemical moieties is determined by applying the moieties to an assay which has been designed to test a particular property of the chemical moiety being screened, for example, a receptor binding assay which tests the ability of the moiety to bind to a particular receptor site.

In an effort to reduce the time and expense involved in screening a large number of randomly chosen compounds for biological activity, several developments have been made to provide libraries of compounds for the discovery of lead compounds. The chemical generation of molecular diversity has become a major tool in the search for novel lead structures. Currently, the known methods for chemically generating large numbers of molecularly diverse compounds generally involve the use of solid phase synthesis, in particular to synthesize and identify peptides and peptide libraries. See, for example, Lebl et al., *Int. J. Pept. Prot. Res.*, **41**, p. 201 (1993) which discloses methodologies providing selectively cleavable linkers between peptide and resin such that a certain amount of peptide can be

liberated from the resin and assayed in soluble form while some of the peptide still remains attached to the resin, where it can be sequenced; Lam et al., *Nature*, **354**, p. 82 (1991) and (WO 92/00091) which disclose a method of synthesis of linear peptides on a solid support such as polystyrene or polyacrylamide resin; Geysen et al., *J. Immunol. Meth.*, **102**, p. 259 (1987) which discloses the synthesis of peptides on derivatized polystyrene pins which are arranged on a block in such a way that they correspond to the arrangement of wells in a 96-well microtiter plate; and Houghten et al., *Nature*, **354**, p. 84 (1991) and WO 92/09300 which disclose an approach to de novo determination of antibody or receptor binding sequences involving soluble peptide pools.

Nonpeptidic organic compounds, such as peptide mimetics, can often surpass peptide ligands in affinity for a certain receptor of enzyme. An effective strategy for rapidly identifying high affinity biological ligands, and ultimately new and important drugs, requires rapid construction and screening of diverse libraries of non-peptidic structures containing a variety of structural units capable of establishing one or more types of interactions with a biological acceptor (e.g., a receptor or enzyme), such as hydrogen bonds, salt bridges, pi-complexation, hydrophobic effects, etc. However, work on the generation and screening of synthetic test compound libraries containing nonpeptidic molecules is now in its infancy. One example from this area is the work of Ellman and Bunin on a combinatorial synthesis of benzodiazepines on a solid support (J. Am. Chem. Soc. 114, 10997, (1992); *see* Chemical and Engineering News, January 18, 1993, page 33).

Historically, the process to identify pharmacologically active agents has been characterized as time-consuming, labor-intensive and inefficient and usually involves a single biological target per screening effort. Further, efforts to improve the process have focused entirely of increasing the number and quality of agents per screening effort. Until now there has been little if any emphasis on improving the efficiency with which biological targets are utilized in screening assays.

Additionally, efforts are underway to fully characterize the human genome. This initiative has already produced thousands of novel gene sequences, many of

which are of unknown function. To identify which of these genes present opportunities for pharmaceutical intervention would take years of diligent effort using current methodology. Presently there is a need for improved methods for efficiently identifying pharmaceutically active agents and for better utilization of information resulting from the human genome initiative.

Many of the disadvantages of the known methods as well as many of the needs not met by them are addressed by the present invention which, as described more fully hereinafter, provides numerous advantages over the known methods.

SUMMARY OF THE INVENTION

This invention relates to a process for identifying pharmaceutically active agents which comprises simultaneously expressing a plurality of Uniquely Tagged Target Agents to form a Target Library; preparing an Agent Candidate Pool and testing the Target Library and the Agent Candidate Pool in an assay which identifies agents of the candidate pool having desired characteristics and to pharmaceutically active agents identified by such process.

This invention also relates to a process for simultaneously expressing a plurality of uniquely tagged genes to form a Target Library.

This invention also relates to a process for simultaneously expressing a plurality of uniquely tagged gene products to form a Target Library.

This invention also relates to a process for simultaneously expressing a plurality of uniquely tagged membrane-bound receptors to form a Target Library.

This invention also relates to a Target Library comprising a plurality of expressed uniquely tagged genes.

This invention also relates to a Target Library comprising a plurality of expressed uniquely tagged gene products.

This invention also relates to a Target Library comprising a plurality of expressed uniquely tagged membrane-bound receptors.

This invention also relates to Spatially Encoded Target Libraries.

This invention also relates to methods of screening for pharmaceutical activity utilizing Spatially Encoded Target Libraries.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the term "Agent Candidate Pool" means a source of one or more entities selected from the group consisting of: chemical compounds, such as mixtures of individual compounds, peptides, natural products and monoclonal antibodies which are to be tested for pharmaceutical activity. These entities are optionally tested as combinatorial compound libraries, combinatorial peptide libraries or variomers.

As used herein, the term "Target Library" means a plurality of expressed Uniquely Tagged Target Agents.

As used herein, the term "Uniquely Tagged Target Agent" means uniquely tagged genes, uniquely tagged gene products or uniquely tagged membrane-bound receptors.

As used herein, the term "gene products" means a protein produced by the expression of a gene. For example, an enzyme, a receptor or a docking protein such as an SH2 domain.

As used herein, the term "uniquely tagged gene(s)" means that, prior to expression, each gene that will make up the Target Library is engineered with a distinct marker such as a short peptide sequence (hereinafter epitope tag) which is capable of recognition by specified antibodies. Said epitope tags can be manufactured in large number. Additionally, several genes can be grouped with the same tag, subgrouped with different tag combinations or tandem tags can be assembled to allow sequential selection.

As used herein, the term "uniquely tagged gene product(s)" means that, prior to expression, each gene product that will make up the Target Library is engineered with a distinct marker such as a short peptide sequence (hereinafter epitope tag) which is capable of recognition by specified antibodies. Said epitope tags can be manufactured in large number. Additionally, several gene products can be grouped with the same tag, subgrouped with different tag combinations or tandem tags can be assembled to allow sequential selection.

As used herein, the term "uniquely tagged membrane-bound receptor(s)" means that, prior to expression, each receptor that will make up the Target Library is engineered with a distinct marker such as a short peptide sequence (hereinafter epitope tag), which is capable of recognition by specified antibodies, and then incorporated into suitable recipient cells. Said epitope tags can be manufactured in large number. Additionally, several receptors can be grouped with the same tag, subgrouped with different tag combinations or tandem tags can be assembled to allow sequential selection. Incorporation of one target receptor type per cell is preferred, as this will enable adjustment of the relative composition of the Target Library by mixing together different numbers of cells. This cell mixture will constitute the Target Library in this example.

Many receptors require a cell membrane for proper binding and functional responses to ligands, for example 7 transmembrane G-protein coupled receptors. As used herein, receptors which require a cell membrane for proper binding and functional response and which are incorporated into a cell are referred to as "membrane-bound receptors". Generally, applying Target Library screening approaches to membrane-bound types of receptors can be performed as indicated herein. A preferred method of screening membrane-bound receptors is to use one or more chemical compounds or a combinatorial chemical library as the Agent Candidate Pool and provide a photoactivatable group in the template during the synthesis of the compounds or combinatorial library. Examples of such groups are aryl azides or benzophenones. Additionally, a readily detectable marker such as biotin is preferably incorporated into the template. Thus, each compound will preferably carry both the photoactivatable and detection groups. Alternatively, if the ligands for the targets are known, the photoactivatable and detection tags need only be incorporated into the ligands, and unlabelled compounds can be screened by competition. In this preferred example of the invention the compound or combinatorial library is mixed with the Target Library and irradiated to activate the photoactivatable group. The cells of the Target Library are lysed and the mixture is added to an array of spatially encoded antibodies raised against the epitope tags of the Target Library. The matrix of antibodies is probed with reagents to detect the

marker (for example: streptavidin-peroxidase to detect biotin). The spectrum of targets that interact with a given compound or group of compounds can be identified by association of the detection marker with a particular subset of epitope tags. If necessary, the Target Library can be pre-screened for incorporation of the detection marker prior to deconvolution on the antibody matrix.

As used herein, the term "simultaneously expressed" means that numerous Uniquely Tagged Target Agents are expressed and purified together without regards to maintaining the identity of the subject gene, gene product or membrane-bound receptor.

As used herein, the term "assay", "screen" or "testing for biological activity" includes any form of testing for pharmaceutically relevant activity.

In preparing the Target Library, one may initially begin with any manageable number of genes, preferably less than one hundred. Preferably these genes are grouped for example by expression in tissues of interest or by membership in a common family, such as kinases, based on sequence homology. Each gene is uniquely tagged preferably with an epitope tag and simultaneously expressed.

In preparing the Target Library, one may initially begin with any manageable number of gene products, preferably less than one hundred. Preferably these genes are grouped for example by expression in tissues of interest or by membership in a common family, such as SH2 domains, based on sequence homology. Each gene product is uniquely tagged preferably with an epitope tag and simultaneously expressed.

In preparing the Target Library, one may initially begin with any manageable number of receptors, preferably less than one hundred. Preferably these receptors are grouped for example by expression in tissues of interest or by membership in a common family, such as 7 transmembrane G-protein coupled receptors, based on sequence homology. Each receptor is uniquely tagged preferably with an epitope tag, incorporated into a cell and simultaneously expressed.

To identify compounds that perturb the function of the targets in the Target Library, an Agent Candidate Pool, preferably one or more chemical compounds or a combinatorial compound library, is prepared by standard methods. Active agents are

identified by known assay techniques or by engineering a selection step based on the anticipated activities contained in the Target Library (as described in the Examples below). Deconvolution of the Target Library can be accomplished by means of the epitope tag, while analytical methods, such as mass spectral techniques, can be used for Agent Candidate Pool analysis. Thus the identification of active agents and the function of the Uniquely Tagged Target Agent are analyzed in parallel.

Pharmaceutically active agents identified by the above process are optionally screened in cell based assays for a function of interest, followed by an optional functional screen or animal model. Also included within the scope to the present invention are pharmaceutically active agents identified by the processes disclosed herein. While is not necessary to identify the specific target (gene, gene product or receptor) affected, if toxicity problems arise this information could be potentially useful and could readily be obtained retrospectively. Further, parallelism is maintained in the secondary screens, with multiple compounds being simultaneously evaluated in a battery of cell-based assays.

A particularly advantageous aspect of this invention is the ability to maintain parallelism in all phases of the discovery process i.e. target selection, expression and purification, compound synthesis, primary and secondary screening. Additional advantages of this process include: the ability for multiple targets to be simultaneously evaluated without prior knowledge of function or disease association, built-in selectivity profiling, greater chance of uncovering totally unexpected activities and disease indications, increased efficiency by utilizing more drug candidates per unit time.

Within the general framework of the parallel approach described above, numerous methodologies for screening Target Libraries are possible. One such methodology is described in Example 1. All such methodologies are within the scope the invention as claimed herein.

In a further aspect of the invention there is provided a preferred method for screening Target Libraries which utilizes expressed Uniquely Tagged Target Agents, wherein the tagging marker is an epitope tag, spatially encoded in discrete

areas of a substrate (referred to herein as Spatially Encoded Target Library or SETL). Spatially Encoded Target Libraries for use herein are prepared by:

i) depositing groupings of antibodies, each grouping being directed against a different epitope tag of the Target Library and each epitope tag of the Target Library corresponding to at least one antibody grouping, in an array (hereinafter antibody region), preferably in a mosaic pattern, on a substrate, preferably a polyvinylidene difluoride (PVDF) membrane, a nitrocellulose membrane, a polystyrene layer or a silicon wafer,

ii) blocking the antibody region of the membrane with an inert protein and

iii) exposing the Target Library to the antibody region.

Antibody groupings will selectively bind to their corresponding epitope tags, thereinby localizing like groups of Uniquely Tagged Target Agents to a specific area of the membrane. The above process is repeated, as desired, in order to produce a plurality of functionally similar SETLs.

The deposition of the antibodies to the membrane can be accomplished by any means, for example manually spotting the membrane with a capillary pipette, or in an automated fashion by 'micro-spotting' in small predetermined deposit areas, for example by the use of printing techniques such as those used in ink jet printers. The membrane carrying the immobilized antibodies is preferably bonded to a vessel, such as Merrifield synthesis vessels, flasks, or preferably microtiter 96-well plates, prior to step ii above, such that each vessel contains an entire antibody region.

The Agent Candidate Pool is portioned into the vessels along with other assay reagents. Following a detection step, the membrane is imaged to identify activity of the Uniquely Tagged Target Agents of the Target Library and the effects of the Agent Candidates.

The SETL approach is advantageously used in developing screening assays for receptors or other binding proteins, for example SH2 domains, as shown in Examples 2, 3 and 4 below.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent.

The following Examples are, therefore, to be construed as merely illustrative and not a limitation of the scope of the present invention in any way.

Assay Examples.

Example 1: Novel osteoblast kinases.

The cell type responsible for new bone production is the osteoblast. As a result of the application of large scale sequencing techniques to osteoblast cDNA libraries, a number of novel full length kinases have been identified. A number of full length kinases from osteoblast-like cells are each uniquely tagged with one or more epitope tags. For example, a unique identifier tag can be attached in tandem with a common purification tag to facilitate the purification process. The genes are tagged and simultaneously expressed in an acceptable host, e.g. E. Coli or baculovirus. The tags are used to simultaneously purify all members of the kinase Target Library as a mixture, which is characterized by Western blotting with antibodies to the unique tags. Suitable substrates are identified by separately immobilizing each member of the Target Library via its epitope tag and adding a mixture of phosphate acceptors along with isotopically labelled ATP. Phosphorylated substrates are identified by mass spectrometric techniques.

A compound combinatorial library is synthesized by standard methods. The Target Library is spatially encoded (as described above) along with suitable substrates. The latter are synthesized so that they contain both a phosphate acceptor and an epitope tag to co-localize them with the relevant kinase targets. Test compounds are added to each array, along with isotopically labelled ATP. The arrays are imaged to identify areas where phosphate has been incorporated.

Active agents from Example 1 are optionally screened in a secondary assay, such as an osteocalcin release assay which correlates with bone formation. Additionally, active agents from Example 1 are screened in a battery of cell based assays covering functions unrelated to bone formation. Preferably, said cell based assay utilizes reporter gene technology, for example - as described in US patent numbers 5,401,629 and 5,436,128. Such technology utilizes the promoter regions of various genes of interest e.g. osteocalcin, coupled to a reporter gene e.g.

luciferase, and transfected into an appropriate cell line. When this is done in a multipotent cell line e.g. embryonic stem cells, the same cells may be used with different reporter constructs to simultaneously assay for a variety of functions (hereinafter 'assay library'). Thus the parallel approach of the present invention is maintained all the way to the secondary assay, giving advantages of efficiency and wide profiling of active compounds.

Preferably the active agents are then evaluated in animal models or in vitro functional assays and screened for toxicity.

Example 2: Binding assay.

The SETL approach could be used to configure a screening assay for receptors or other binding proteins, such as SH2 domains. SH2 domains bind phosphorylated peptide ligands. To configure a SETL assay for SH2 domains, the domains are expressed as epitope tagged fusion proteins. Antibodies to the tags are deposited on membranes as above. A mixture of tagged SH2 domains is added to each well and allowed to bind to the corresponding antibodies. A mixture of biotinylated phosphopeptide ligands (one for each SH2 target) is added along with test compounds. Binding of the biotinylated peptide ligands is detected by addition of streptavidin-biotin complex and colorimetric, fluorescence or enhanced chemiluminescence (ECL) detection reagents. Imaging of the activity of the detection reagents on the membrane surface allows binding of the ligand to each individual SH2 domain to be measured. Such imaging could be automated.

A working demonstration of this assay was accomplished by expressing the human src SH2 domain as a fusion protein.

The fusion protein containing the human src SH2 domain was expressed as the general sequence: DET1-DET2-spacer-ek-src SH2, where DET1, DET2, spacer and ek are as described below. DET1 ("defined epitope tag 1") (SEQ ID NO: 1) is an 11 amino acid sequence found in the Human Immunodeficiency Virus Type 1 (HIV-1) envelope protein gp120 (or gp160). Monoclonal antibodies to various epitopes of HIV-1 gp120 (or gp160) are known in the art, see, for example U.S. Patent 5,166,050. One preferred example is monoclonal antibody 178.1 (see, e.g.,

Thiriart et al., J. Immunol., 143:1832-1836 (1989)), which was prepared by immunization of mice with a yeast-expressed HIV-1 gp160 molecule from strain BH10 (Ratner et al., Nature, 313:277-284 (1985)). This tag was used for detection of expression (by Western blot), for purification of the protein (by affinity chromatography), and for configuring assays in which the fusion protein was captured or immobilized using the 178.1 antibody. DET2 is a hexa-histidine sequence tag (SEQ ID NO: 2) which binds to nickel-containing resins and was used for purification purposes. Spacer (SEQ ID NO: 3) was utilized to design a BamH1 restriction site at the indicated position of the construct. The term -ek- refers to a recognition sequence (SEQ ID NO: 4) for the enterokinase protease which provides for the optional removal of the tags from the src SH2 domain, thus producing a src SH2 domain that contains no extraneous amino acids. A src SH2 domain which contains no extraneous amino acids are preferable to tagged protein for crystallography studies.

The DNA sequence encoding each DET1-DET2-spacer-ek-src SH2 was designed such that the indicated restriction sites (BamH1 and XbaI) flank the spacer-ek-src SH2 region, thereby allowing different spacer-ek-SH2 contructs to be readily substituted into any one of the vectors as described in Procedure 2 below to create a DET1-DET2-spacer-ek-SH2 tagged protein. The DNA sequence encoding the DET1-DET2-spacer-ek-src SH2 construct was also designed such that the entire tagged SH2 domain can be moved as an NdeI-XbaI fragment into any expression vector containing an NdeI site at an appropriate distance downstream of E. coli transcription and translation regulatory sequences and a downstream cloning site compatible with XbaI. Although any suitable vector would yield similar results(e.g., pET-11a; Novagen, Inc.), the vector used in the instant experiments was E. coli expression vector pEA1KnRBS3. This vector is a derivative of the series of vectors described in Shatzman, A, Gross, M, and Rosenberg, M, 1990, "Expression using vectors with phage lambda regulatory sequences", In: Current Protocols in Molecular Biology (F.A. Ausubel et al , eds.), pp. 16.3.1-16.3.11, Greene Publishing and Wiley-Interscience, N.Y. (hereinafter F.A. Ausubel et al.). The specific vector pEA1KnRBS3 is described in Bergsma et al, 1991, J. Biol. Chem. 266:23204-23214.

The procedures below describe the expression of chicken src and the human src SH2 domains. First, the chicken src SH2 domain was expressed as DET1-DET2-spacer-SH2. Then, the other was inserted into this vector in place of chicken src to express proteins in the form DET1-DET2-spacer-ek-src SH2 as described in procedures 1 and 2 below.

Procedure 1: Cloning and Expression of chicken src SH2 domain containing tags DET1 and DET2 (DET1-DET2-spacer-SH2).

A DNA sequence encoding the tagged protein DET1-DET2-spacer-SH2 was PCR amplified from a cDNA clone containing the chicken src gene (p5H; Levy et al 1986. Proc. Natl. Acad. Sci. USA 83:4228) by methods well known to those skilled in the art by using the following primers:

5'

TTCCATATGAAAAGTATTCGTATTCAGCGTGGCCCGGGCCGTCACCACCA CCACCACCACGGGATCCCCGCTGAAGAGTGGTACTTT 3' (SEQ ID NO: 7)

The underlined sites are an NdeI recognition site (5') and a BamHI recognition site (3').

5' GGAATTCTAGATTACTAGGACGTGGGGCAGACGTT 3' (SEQ ID NO: 8)

The underlined region is an XbaI recognition site.

The PCR product was digested with NdeI and XbaI, followed by isolation of the digested fragment on an agarose gel. The fragment was ligated into NdeI-XbaI-digested pEA1KnRBS3 vector (Bergsma et al, supra) that had been agarose gel purified as a 6.5 kbp fragment. The ligation reaction was used to transform E. coli MM294cI$^+$ (F.A. Ausubel et al., supra). A plasmid containing an insertion of the correct fragment was identified and confirmed by DNA sequencing. The resultant plasmid encodes DET1-DET2-spacer-SH2 under the control of the phage lamda P_L promoter and regulatory system. Plasmid DNA was purified from MM294cI$^+$ and

used to transform E. coli strain AR120. In this host strain, expression of the phage promoter can be induced by addition of nalidixic acid to the growing culture as described in F.A. Ausubel et al, supra. Nalidixic acid induction of AR120 containing this plasmid, followed by analysis of the cellular proteins on an SDS-polyacrylamide gel stained with Coomassie Blue (F.A. Ausubel et al., supra), resulted in appearance of a protein band with an apparent molecular weight of 15,000; this band was not seen in uninduced cells or in induced cells containing pEA1KnRBS3 lacking the PCR amplified fragment. Western blotting confirmed that the induced protein band reacted with the anti-DET1 monoclonal antibody 178.1.

Procedure 2: Cloning, expression and purification of human src SH2 domain containing tags and an enterokinase proteolytic cleavage site (DET1-DET2-spacer-ek-src SH2).

A DNA sequence encoding protein ek-src SH2 was PCR amplified from a cDNA clone containing the human src gene (c-src SH2 DNA sequence identical to that described in Takeya,T. and Hanafusa, H, 1983 Cell 32:881-890) using the following primers:

5' CGGGATCCTGGACGACGACGACAAAGCTGAGGAGTGGTATTTT 3' (SEQ ID NO: 9)

The underlined site is a BamHI recognition site.

5' GGAATTCTAGACTATTAGGACGTGGGGCACACGGT 3' (SEQ ID NO: 10)

The underlined region is an XbaI recognition site.

The PCR product was digested with BamHI and XbaI, followed by isolation of the digested fragment on an agarose gel. The fragment was ligated into BamHI-XbaI-digested expression vector containing the tagged chicken src gene DET1-

DET2-spacer-SH2 described in Procedure 1 above. In that vector, the BamHI site is located between the coding regions for DET2 and SH2, and the XbaI site is located after the 3' end of the SH2 coding region. The ligation reaction was used to transform E. coli MM294cI$^+$. The construct DET1-DET2-spacer-ek-src SH2 was confirmed by DNA sequencing (SEQ ID NO: 5) and induced in E. coli strain AR120 as described in Procedure 1 above. A Coomassie-Blue-stained, Western-blot-positive induced protein band with an apparent molecular weight of 16,000 was observed after nalidixic acid induction.

Cells were lysed at neutral pH by sonication in the presance of lysozyme. After centrifugation, the soluble extract was chromatographed on a Ni^{++}NTA column. After washing the column with equilibration buffer (Tris buffer pH 8 containing 0.5 M NaCl) and the same buffer containing 15 mM imidazole, the protein, DET1-DET2-spacer-ek-src SH2, was eluted in highly purified form with 25 mM imidazole in equilibration buffer.

After manually spotting Mab 178.1 on a PVDF membrane and addition of tagged src SH2 domain, probing with a biotinylated phosphopeptide ligand for src SH2 (SEQ ID NO: 6), followed by ECL or colorimetric detection showed a clear signal. No signal was seen with an equivalent amount of irrelevant antibody. Binding of the biotinylated phosphopeptide to the immobilized src SH2 domain was inhibited by addition of a non-biotinylated phosphopeptide competitor of identical amino acid sequence to the biotinylated molecule.

Example 3: Protease assay.

As with the kinase assay, it is necessary to identify the optimum substrates for the target library components before configuring the SETL screen. Once this is known, substrates are synthesized with the appropriate tag on one side of the cleavage site and a detection label such as biotin on the other. Once immobilized, protease activity can be identified by loss of the label from the target site.

Example 4: Biopanning method

Screening of bead-based compound libraries with multiple targets.

Each gene of a Target Library is expressed with two epitope tags: a common tag C which is the same for all genes and a unique tag T which is different for each gene. A combinatorial library is synthesized on beads by standard methods. The combinatorial library is screened with the Target Library by measurement of the association of tag C with the beads and separation of individual positive beads either manually or by automated methods. Following separation, the proteins are eluted from each positive bead, for example by low pH washing, and allowed to bind to an array of spatially encoded antibodies against the unique tags T. The location, thus identification, of the spectrum of proteins from the positive beads is determined by probing the matrix with an antibody against C, such probe being labeled with a readily detectable marker such as biotin, with subsequent readout. Further, active compounds are identified by cleavage from the positive beads with subsequent chemical/physical analysis.

While the preferred embodiments of the invention are illustrated by the above, it is to be understood that the invention is not limited to the precise instructions herein disclosed and that the right to all modifications coming within the scope of the following claims is reserved.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: DUNNINGTON, DAMIEN

(ii) TITLE OF THE INVENTION: A PROCESS FOR IDENTIFYING PHARMACEUTICALLY ACTIVE AGENTS

(iii) NUMBER OF SEQUENCES: 10

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: SmithKline Beecham Corporation
(B) STREET: 709 Swedeland Road
(C) CITY: King of Prussia
(D) STATE: PA
(E) COUNTRY: USA
(F) ZIP: 19406

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ Version 1.5

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: 08/497,357
(B) FILING DATE: 30-Jun-1995

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Dustman, Wayne J
(B) REGISTRATION NUMBER: 33,870
(C) REFERENCE/DOCKET NUMBER: P50323-3

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 610-270-5023
(B) TELEFAX:
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 11 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: internal
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Lys Ser Ile Arg Ile Gln Arg Gly Pro Gly Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

His His His His His His
1 5

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(iii) HYPOTHETICAL: NO

(iv) ANTISENSE: NO

(v) FRAGMENT TYPE: internal

(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Gly Ile Leu
1

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: internal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Asp Asp Asp Asp Lys
1 5

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 130 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: internal
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met Lys Ser Ile Arg Ile Gln Arg Gly Pro Gly Arg His His His His
1 5 10 15
His His Gly Ile Leu Asp Asp Asp Asp Lys Ala Glu Glu Trp Tyr Phe
20 25 30
Gly Lys Ile Thr Arg Arg Glu Ser Glu Arg Leu Leu Leu Asn Ala Glu
35 40 45
Asn Pro Arg Gly Thr Phe Leu Val Arg Glu Ser Glu Thr Thr Lys Gly
50 55 60

```
Ala Tyr Cys Leu Ser Val Ser Asp Phe Asp Asn Ala Lys Gly Leu Asn
65                  70                  75                  80
Val Lys His Tyr Lys Ile Arg Lys Leu Asp Ser Gly Gly Phe Tyr Ile
                85                  90                  95
Thr Ser Arg Thr Gln Phe Asn Ser Leu Gln Gln Leu Val Ala Tyr Tyr
            100                 105                 110
Ser Lys His Ala Asp Gly Leu Cys His Arg Leu Thr Thr Val Cys Pro
        115                 120                 125
Thr Ser
    130
```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 19 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE: internal
(vi) ORIGINAL SOURCE:
(ix) FEATURE:

(A) NAME/KEY: Other
(B) LOCATION: 4...4
(D) OTHER INFORMATION: phosphorylated tyrosine r
esidue

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```
Glu Pro Gln  Tyr Glu Glu Ile Pro Ile Tyr Leu
 1              5                  10                  15
```

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 87 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TTCCATATGA AAAGTATTCG TATTCAGCGT GGCCCGGGCC GTCACCACCA CCACCACCAC
60
GGGATCCCCG CTGAAGAGTG GTACTTT
87

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GGAATTCTAG ATTACTAGGA CGTGGGGCAG ACGTT
38

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 46 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CGGGATCCTG GACGACGACG ACAAAGCTGA GGAGTGGTAT TTT
46

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 38 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: cDNA
(iii) HYPOTHETICAL: NO
(iv) ANTISENSE: NO
(v) FRAGMENT TYPE:
(vi) ORIGINAL SOURCE:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGAATTCTAG ACTATTAGGA CGTGGGGCAC ACGGT

38

5

What is claimed is:

1. A process for identifying pharmaceutically active agents which comprises the steps of:

a) simultaneously expressing a plurality of Uniquely Tagged Target Agents to form a Target Library;

b) preparing an Agent Candidate Pool; and

c) testing the Target Library and the Agent Candidate Pool in an assay which identifies agents having desired characteristics.

2. A process for spatially encoding Target libraries which comprises the steps of:

a) depositing a plurality of antibodies, each member being directed against a different epitope tag of a Target Library and each epitope tag of the Target Library corresponding to at least one antibody grouping, in an array, on a substrate;

b) blocking the antibody region of the membrane with an inert protein; and

c) exposing a Target Library to the antibody region.

3. A spatially Encoded Target Library prepared as in claim 2.

4. A process for simultaneously expressing a plurality of uniquely tagged genes to form a Target Library which comprises engineering each member of the plurality with an epitope tag capable of being recognized by a specified antibody, prior to simultaneously expressing the plurality.

5. A Target Library prepared as in claim 4.

6. A method of screening bead based combinatorial compound libraries which comprises the steps of:

a) expressing a Target Library wherein to each member is affixed a common tag and a unique tag;

b) exposing a bead based combinatorial compound library to the Target Library and separating the positive hits by detection of the common tag;

c) eluting the active proteins off of the beads; and

d) identification by means of the unique tag.

5

7. Pharmaceutically active agents identified by the process of claim 1.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US96/02490

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) :G01N 33/53
US CL :435/6, 7.1; 536/518
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 7.1; 536/518

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, DIALOG
search terms: library, phage-display, epitope tag, combinatorial library

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DAVIDSON ET AL. "Folded proteins occur frequently in libraries of random amino acid sequences". Proceedings of the National Academy of Sciences. March 1994, Vol. 91, pages 2146-2150, see Abstract and Figure 1.	1, 4, 5
X, E --- Y, E	US 5,498,538 A (KAY ET AL.) 12 March 1996, columns 7 and 21.	1, 4-7 --- 2, 3
Y, P	US 5,482,867 A (BARRETT ET AL.) 09 January 1996, columns 2-3.	2, 3

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be part of particular relevance	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E"	earlier document published on or after the international filing date	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&"	document member of the same patent family
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search
22 MAY 1996

Date of mailing of the international search report
27 JUN 1996

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer
LORA M. GREEN
Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)*